U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's
Name: _____

Serial
Number: _____

Date: _____   Phone: _____   Art Unit: _____

Search Topic:
Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevent citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevent claim(s).

112526

Shears, Beverly

From: Devi, Sarvamangala
Sent: Monday, January 19, 2004 11:45 AM
To: Shears, Beverly
Subject: 09/738,599

Beverly:
        Please perform a sequence search for the following in application SN 09/738,599:

        A single nucleic acid molecule comprising 1 to 33 nucleotides of SEQ ID NO: 21 located 5' to nucleotides 73 to 309 of SEQ ID NO: 22, or 5' to a 12 nucleotide-long fragment of 73 to 309 of SEQ ID NO: 22.

Thanks.

S. DEVI, Ph.D.

STIC Search Report
Biotech-Chem Library

Scientific & Technical Information Center
S T I C
USPTO

STIC Database Tracking Number: 112526

To: Sarvamangala Devi
Location: CM1/7E15/7E12
Art Unit: 1645
Wednesday, January 21, 2004

Case Serial Number: 09/738599

From: Beverly Shears
Location: Remsen Bldg.
            RM 1A54
Phone: 571-272-2528

beverly.shears@uspto.gov

Search Notes

STAFF USE ONLY

Date completed: _____

Searcher: __Beverly e 2528__

Terminal time: _____

Elapsed time: _____

CPU time: _____

Total time: _____

Number of Searches: _____

Number of Databases: ___1___

Search Site
____ STIC
____ CM-1
____ Pre-S

Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
__✓__ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
__✓__ Other CGN

PTO-1590 (9-90)

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 22:54:00 ; Search time 1535 Seconds
                                      (without alignments)
                                      7195.830 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-09-738-599-22
Perfect score: 270
Sequence: 1 ctggaagtctgtccaggg..............GGGTATATTGCTCACAATAG 270

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
  M
  E

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422
Maximum DB seq length: 200000000
Minimum DB seq length: 0
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*

```
 1: gb_ba:*
 2: gb_htg:*
 3: gb_in:*
 4: gb_om:*
 5: gb_ov:*
 6: gb_pat:*
 7: gb_ph:*
 8: gb_pl:*
 9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_pr:*
27: em_ro:*
28: em_sts:*
29: em_un:*
30: em_vi:*
31: em_htg_hum:*
32: em_htg_inv:*
33: em_htg_other:*
34: em_htg_mus:*
35: em_htg_pln:*
36: em_htg_rod:*
37: em_htg_mam:*
38: em_htg_vrt:*
39: em_sy:*
40: em_hgo_hum:*
41: em_hgo_mus:*
```

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	237	87.8	309	6	AR102880	AR102880 Sequence
2	237	87.8	309	6	AR129424	AR129424 Sequence
3	237	87.8	378	6	AR102879	AR102879 Sequence
4	237	87.8	378	1	AR129423	AR129423 Sequence
5	237	87.8	760	1	AP042279	AP042279 Escherich
6	237	87.8	760	6	AR102863	AR102863 Sequence
7	237	87.8	760	6	AR129407	AR129407 Sequence
8	237	87.8	31869	1	AY205565	AY205565 Escherich
9	210.4	77.9	306358	1	AB016759	AB016759 Escherich
10	205.6	76.1	309	6	AR102867	AR102867 Sequence
11	205.6	76.1	309	6	AR129411	AR129411 Sequence
12	205.6	76.1	326	7	BLAMLOM	X55792 Bacteriopha
13	205.6	76.1	9170	12	U02453	U02453 Cloning vec
14	205.6	76.1	9205	12	XXU02427	U02427 Cloning vec
15	205.6	76.1	11313	1	AB005297	AB005297 Escherich
16	205.6	76.1	11933	6	AR138377	AR138377 Sequence
17	205.6	76.1	11933	6	AR146887	AR146887 Sequence
18	205.6	76.1	42529	12	CVI39284	U39284 Cloning vec
19	205.6	76.1	42530	12	CVI39285	U39285 Cloning vec
20	205.6	76.1	42531	12	CVI39286	U39286 Cloning vec
21	205.6	76.1	42704	12	CVI37692	U37692 Cloning vec
22	205.6	76.1	48502	7	LAMCG	J02459 Bacteriopha
23	205.6	76.1	61663	7	AR204166	AR204166 Sequence
24	205.6	76.1	61670	7	AF125520	AF125520 Bacteriop
25	205.6	76.1	61765	7	AP004402	AP004402 Stx2 conv
26	205.6	76.1	65510	6	AP000422	AP000422 Escherich
27	205.6	76.1	327773	1	AP002554	AP002554 Escherich
28	205.6	75.0	14164	1	AB005330	AB005330 Escherich
29	202.4	75.0	46819	6	AR204176	AR204176 Sequence
30	202.4	75.0	222605	1	AP002555	AP002555 Escherich
31	201.2	74.5	309	6	AR102866	AR102866 Sequence
32	201.2	74.5	309	6	AR129410	AR129410 Sequence
33	201.2	74.5	1430	1	ECISSGENE	ECISSGENE
34	200.8	74.4	16170	1	AB000161	AB000161 Escherich
35	200.8	74.4	136742	1	ECU82598	U82598 Escherichia
36	200.8	74.4	250010	2	AC074027	AC074027 Mus muscu
37	141.2	52.3	752	1	AP441251	AP441251 Paracocci
38	71.2	26.4	180	1	AF222296	AF222296 Escherich
39	71.2	26.4	180	6	AX275633	AX275633 Sequence
40	41.6	15.4	164943	2	AC111660	AC111660 Rattus no
41	41.6	15.4	173268	2	AC112087	AC112087 Rattus no
42	41	15.2	171053	2	AC113499	AC113499 Mus muscu
43	40.8	15.1	822	8	AX510607	AX510607 Sequence
44	40.8	15.1	97451	8	F19K23	AC000375 Sequence
45	39.2	14.5	236022	2	AC105604	AC105604 Rattus no

ALIGNMENTS

```
RESULT 1
AR102880
LOCUS       AR102880      309 bp    DNA    linear    PAT 14-FEB-2001
DEFINITION  Sequence 22 from patent US 6087128.
ACCESSION   AR102880
VERSION     AR102880.1  GI:12814468
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Nolan,L.K. and Horne,S.M.
  TITLE     DNA encoding an avian E. coli iss
  JOURNAL   Patent: US 6087128-A 22 11-JUL-2000;
FEATURES             Location/Qualifiers
```

Pred. No. is the number of results predicted by chance to have a

```
BASE COUNT      101 a      60 c      64 g      84 t
ORIGIN
source          1...309
                /organism="unknown"

RESULT 2
AR129424
LOCUS       AR129424              309 bp    DNA       linear    PAT 16-MAY-2001
DEFINITION  Sequence 22 from patent US 6187321.
ACCESSION   AR129424
VERSION     AR129424.1  GI:14117321
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Nolan,L.K., Horne,S.M. and Robinson,M.
  TITLE     Avian E. coli 188 polypeptide
  JOURNAL   Patent: US 6187321-A 22 13-FEB-2001;
FEATURES             Location/Qualifiers
  source             1..309
                     /organism="unknown"
BASE COUNT      101 a      60 c      64 g      84 t
ORIGIN

Query Match    87.8%;  Score 237;  DB 6;  Length 309;
Best Local Similarity  100.0%;  Pred. No. 2.4e-52;
Matches 237;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   34  CAAACGTTTACTGTTGGAAACAAACCGACAGCAGTAACACCAAAGGAAAACCATCACTCAT  93
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73  CAAACGTTTACTGTTGGAAACAAACCGACAGCAGTAACACCAAAGGAAAACCATCACTCAT  132

Qy   94  CATTTCTTCGTTTCGGGACAAGAGAAAAACTGTTGATGCAGCCAAAAATTTGTGTGGC  153
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133  CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAAACTGTTGATGCAGCCAAAAATTTGTGTGGC  192

Qy  154  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  213
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  252

Qy  214  TTTATCACTTTTGGGCATCTATACTCCGCTGGAAGCCCGGGGTATATTGCTCACAATAG  270
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253  TTTATCACTTTTGGGCATCTATACTCCGCTGGAAGCCCGGGGTATATTGCTCACAATAG  309

RESULT 3
AR102879
LOCUS       AR102879              378 bp    DNA       linear    PAT 14-FEB-2001
DEFINITION  Sequence 21 from patent US 6087128.
ACCESSION   AR102879
VERSION     AR102879.1  GI:12814467
```

```
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   Nolan,L.K. and Horne,S.M.
  TITLE     DNA encoding an avian E. coli 188
  JOURNAL   Patent: US 6087128-A 21 11-JUL-2000;
FEATURES             Location/Qualifiers
  source             1..378
                     /organism="unknown"
BASE COUNT      112 a      80 c      84 g      102 t
ORIGIN

RESULT 4
AR129423
LOCUS       AR129423              378 bp    DNA       linear    PAT 16-MAY-2001
DEFINITION  Sequence 21 from patent US 6187321.
ACCESSION   AR129423
VERSION     AR129423.1  GI:14117320
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   Nolan,L.K., Horne,S.M. and Robinson,M.
  TITLE     Avian E. coli 188 polypeptide
  JOURNAL   Patent: US 6187321-A 21 13-FEB-2001;
FEATURES             Location/Qualifiers
  source             1..378
                     /organism="unknown"
BASE COUNT      112 a      80 c      84 g      102 t
ORIGIN

Query Match    87.8%;  Score 237;  DB 6;  Length 378;
Best Local Similarity  100.0%;  Pred. No. 2.4e-52;
Matches 237;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   34  CAAACGTTTACTGTTGGAAACAAACCGACAGCAGTAACACCAAAGGAAAACCATCACTCAT  93
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  106  CAAACGTTTACTGTTGGAAACAAACCGACAGCAGTAACACCAAAGGAAAACCATCACTCAT  165

Qy   94  CATTTCTTCGTTTCGGGACAAGAGAAAAACTGTTGATGCAGCCAAAAATTTGTGTGGC  153
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  166  CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAAACTGTTGATGCAGCCAAAAATTTGTGTGGC  225

Qy  154  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  213
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  226  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  285

Qy  214  TTTATCACTTTTGGGCATCTATACTCCGCTGGAAGCCCGGGGTATATTGCTCACAATAG  270
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  286  TTTATCACTTTTGGGCATCTATACTCCGCTGGAAGCCCGGGGTATATTGCTCACAATAG  342
```

```
Query Match    87.8%;  Score 237;  DB 6;  Length 378;
Best Local Similarity  100.0%;  Pred. No. 2.4e-52;
Matches 237;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   34  CAAACGTTTACTGTTGGAAACAAACCGACAGCAGTAACACCAAAGGAAAACCATCACTCAT  93
Db  106  CAAACGTTTACTGTTGGAAACAAACCGACAGCAGTAACACCAAAGGAAAACCATCACTCAT  165

Qy   94  CATTTCTTCGTTTCGGGACAAGAGAAAAACTGTTGATGCAGCCAAAAATTTGTGTGGC  153
Db  166  CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAAACTGTTGATGCAGCCAAAAATTTGTGTGGC  225

Qy  154  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  213
Db  226  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  285

Qy  214  TTTATCACTTTTGGGCATCTATACTCCGCTGGAAGCCCGGGGTATATTGCTCACAATAG  270
Db  226  GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAAAACATTCGTAAAATGGATTGCTCGGT  285
```

```
Db   286 TTTATCACTTTTGGCATCTATACTCCGCTGGGAAGCCCGGGTATATATTGCTCACAATAG 342
```

RESULT 5

```
LOCUS        AF042279                  760 bp    DNA     linear   BCT 18-SEP-2001
DEFINITION   Escherichia coli Iss (iss) gene, complete cds.
ACCESSION    AF042279
VERSION      AF042279.1  GI:5305229
KEYWORDS     .
SOURCE       Escherichia coli
  ORGANISM   Escherichia coli
             Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
             Enterobacteriaceae; Escherichia.
REFERENCE    1  (bases 1 to 760)
  AUTHORS    Horne,S.M., Pfaff-McDonough,S.J., Giddings,C.W. and Nolan,L.K.
  TITLE      Cloning and sequencing of the iss gene from a virulent avian
             Escherichia coli
  JOURNAL    Avian Dis. 44 (1), 179-184 (2000)
   MEDLINE   20199920
   PUBMED    10737659
REFERENCE    2  (bases 1 to 760)
  AUTHORS    Horne,S.M., Nolan,L.K., Giddings,C.W. and Pfaff,S.J.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-JAN-1998) Vet. & Micro. Sci., North Dakota State
             Univ., Van.Es Hall, Fargo, ND 58105, USA
FEATURES             Location/Qualifiers
     source          1..760
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="102"
                     /db_xref="taxon:562"
                     /note="may be on a large R plasmid"
     gene            292..600
                     /gene="iss"
     CDS             292..600
                     /gene="iss"
                     /note="similar to Bacteriophage lambda Bor protein, and to
                     iss protein from a septicemic human Escherichia coli
                     isolate; causes avian colibacillosis"
                     /codon_start=1
                     /transl_table=11
                     /product="iss"
                     /protein_id="AAD41540.1"
                     /db_xref="GI:5305230"
                     /translation="MQDNKMKKMLFSAALAMLITGCAQQTFVVGNKPTAVPKETITH
                     HFFVSGIGQEKTVDAARICGGABNVVKTETQQTFVNGLLGFITFGITTPLERAVYCSQ
                     "

BASE COUNT      216 a    159 c    160 g    225 t
ORIGIN

Query Match
Best Local Similarity   87.8%;   Score 237;   DB 1;   Length 760;
Matches  237;   Conservative  0;   Mismatches  0;   Indels   0;   Gaps   0;

Qy    34 CAAACGTTTACTGTTGGAAACAAACCGACACAGCAGCAGTAACACTGTTGATGTGGC 93
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   364 CAAACGTTTACTGTTGGAAACAAACCGACAGCAGCAGTAACACTGTTGATGTGGC 423

Qy    94 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAACTGTTGATGCAGCCAAAAATTTGTTGGC 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAACTGTTGATGCAGCCAAAAATTTGTTGGC 483

Qy   154 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAAACATTCGTAAATGGATTGCTCGGT 213
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAAACATTCGTAAATGGATTGCTCGGT 543

Qy   214 TTTATCACTTTTGGCATCTATACTCCGCTGGAAGCCCGGGTATATATTGCTCACAATAG 270
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TTTATCACTTTTGGCATCTATACTCCGCTGGGAAGCCCGGGTATATATTGCTCACAATAG 600
```

RESULT 6

```
LOCUS        AR102863                  760 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION   Sequence 1 from patent US 6087128.
ACCESSION    AR102863
VERSION      AR102863.1  GI:12814451
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 760)
  AUTHORS    Nolan,L.K. and Horne,S.M.
  TITLE      DNA encoding an avian E. coli iss
  JOURNAL    Patent: US 6087128-A 1 11-JUL-2000;
FEATURES             Location/Qualifiers
     source          1..760
                     /organism="unknown"
BASE COUNT      216 a    159 c    160 g    225 t
ORIGIN

Query Match
Best Local Similarity   87.8%;   Score 237;   Pred. No. 2.4e-52;
Matches  237;   Conservative  0;   Mismatches  0;   Indels   0;   Gaps   0;

Qy    34 CAAACGTTTACTGTTGGAAACAAACCGACACAGCAGCAGTAACACTGTTGATGTGGC 93
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   364 CAAACGTTTACTGTTGGAAACAAACCGACAGCAGCAGTAACACTGTTGATGTGGC 423

Qy    94 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAACTGTTGATGCAGCCAAAAATTTGTTGGC 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAACTGTTGATGCAGCCAAAAATTTGTTGGC 483

Qy   154 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAAACATTCGTAAATGGATTGCTCGGT 213
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAAACATTCGTAAATGGATTGCTCGGT 543

Qy   214 TTTATCACTTTTGGCATCTATACTCCGCTGGAAGCCCGGGTATATATTGCTCACAATAG 270
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TTTATCACTTTTGGCATCTATACTCCGCTGGGAAGCCCGGGTATATATTGCTCACAATAG 600
```

RESULT 7

```
LOCUS        AR129407                  760 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION   Sequence 1 from patent US 6187321.
ACCESSION    AR129407
VERSION      AR129407.1  GI:14117304
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 760)
  AUTHORS    Nolan,L.K., Horne,S.M. and Robinson,M.
  TITLE      Avian E. coli iss polypeptide
  JOURNAL    Patent: US 6187321-A 13-FEB-2001;
FEATURES             Location/Qualifiers
     source          1..760
                     /organism="unknown"
BASE COUNT      216 a    159 c    160 g    225 t
ORIGIN

Query Match
Best Local Similarity   87.8%;   Score 237;   DB 6;   Length 760;
Matches  237;   Conservative  0;   Mismatches  0;   Indels   0;   Gaps   0;

Qy    34 CAAACGTTTACTGTTGGAAACAAACCGACACAGCAGCAGTAACACTGTTGATGTGGC 93
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   364 CAAACGTTTACTGTTGGAAACAAACCGACAGCAGCAGTAACACTGTTGATGTGGC 423

Qy    94 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAACTGTTGATGCAGCCAAAAATTTGTTGGC 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAACTGTTGATGCAGCCAAAAATTTGTTGGC 483

Qy   154 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAAACATTCGTAAATGGATTGCTCGGT 213
```

```
RESULT 8
Db  AY205565/c
        484 GGTGCAGAAAAAGTTGTGTTAAAAACAGAAAACTCAGAACAACATTCGTATAAATGGATTGCTCACAATAG 543
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  214 ...                                                                   270
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  544 TTTATCACTTTTGGCATCTATACTCCCGCTGGAAGCCCGGGTATATTGCTCACAATAG          600
```

LOCUS   AY205565   31869 bp   DNA   linear   BCT 26-MAY-2003
DEFINITION Escherichia coli plasmid p300 iro gene cluster, complete sequence;
            and unknown genes.
ACCESSION   AY205565
VERSION   AY205565.1 GI:31075350
KEYWORDS   .
SOURCE   Escherichia coli
ORGANISM   Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1 (bases 1 to 31869)
AUTHORS   Sorsa,L.J., Dufke,S., Heesemann,J. and Schubert,S.
TITLE   Characterization of an iroBCDEN Gene Cluster on a Transmissible
            Plasmid of Uropathogenic Escherichia coli: Evidence for Horizontal
            Transfer of a Chromosomal Virulence Factor
JOURNAL   Infect. Immun. 71 (6), 3285-3293 (2003)
MEDLINE   22646068
PUBMED   12761110
REFERENCE   2 (bases 1 to 31869)
AUTHORS   Schubert,S., Sorsa,J.L., Dufke,S. and Heesemann,J.
TITLE   Direct Submission
JOURNAL   Submitted (20-DEC-2002) Bacteriology, Max von Pettenkofer-Institut,
            Pettenkoferstr. 9a, Munich 80336, Germany
FEATURES   Location/Qualifiers
     source   1..31869
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="562"
                     /db_xref="taxon:562"
                     /plasmid="p300"
     CDS   3..413
                     /note="Orf1; similar to TnpA transposase of transposon
                     Tn1721, Swiss-Prot Accession Number P51565; contains
                     conserved domain similar to Escherichia coli transposases
                     Tn3, Tn1, Tn1721, Tn2501 and Tn3926 (pfam01526.6,
                     transposase)"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42480.1"
                     /db_xref="GI:31075361"
                     /translation="MGRIERTLFLFLIDWLIQSVELRPPRACRLNKGBARNSLARAVFPNR
                     LGIRDRSPEQORYRASGLNLVTANIVLWNTVLERATOGLVERAGKPVDGELLQFLSP
                     LGWEHINLTGDYWRQSRRLBDGKFRPLRMPGKP"
     repeat_region   409..446
                     /note="IRR II of Tn1721"
                     /rpt_type=inverted
     CDS   complement(468..791)
                     /note="Orf2; similar to Salmonella typhimurium plasmid pR64
                     YcjA, GenBank Accession number BAB91595; contains
                     conserved helix-turn-helix motive similar to bacterial
                     repressor proteins (smart00418, pfam01022)"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42481.1"
                     /db_xref="GI:31075362"
                     /translation="MQLEEVAKALKELGHPTRLPIFKHLVKAGBQGLPVGELQKOLGI
                     PSSTLSHHISALVSGLVTQNRESRTLMCVSQYBILRAIIBFLRBBCCVNSKTDVABP
                     AGKNG"
     CDS   896..2029
                     /note="Orf3; similar to Salmonella typhimurium plasmid R64
                     Ydaa, GenBank Accession Number BAB91596; contains two
                     conserved domains of the family of integral membrane
                     proteins predicted to be permeases of unknown specificity
                     (pfam03773.2, DUF318)"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42482.1"
                     /db_xref="GI:31075363"
                     /translation="MNNSWIPMLQDPAAEMPVPLAVERLSLLPIVISAGVSLIRQKVPDHK
                     IQOMMGAARKGKGYLLLASLGAVTPFCCSGTSTIPMLRGLLSLSAKAGFGPTLTLFLVSPLLN
                     PIIVGLMWTFGWKVTLLVAILIAAGVSVLSSIILDYLGFERHIVEYKNSVSGSCATKC
                     GDSEASVKTSAVBSCCGAGLALASVKTNCCTSSAXTIINLKTVKKEGONISACCPSILS
                     EKSSEBSCCSESBSQGNRNLTMNATSGLIKLAMRDALQGFRKDYLPVLILLSVIGSPIYGF
                     IPSEWIAAHAGADNPLAIPLSAVVGIPLVIRABAVIPLASVLMTKGMGLGALMALIIG
                     SAGASLTEVILLKSMFRMPMIIAFITVILGMAILMGYLTQFLP"
     CDS   complement(2322..2939)
                     /note="Orf4; similar to Salmonella typhimurium plasmid
                     pR64 ydbA, GenBank Accession Number BAB91597"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42483.1"
                     /db_xref="GI:31075364"
                     /translation="MGKYQBEKKTPIIMVKRRTLLPPSLSEKTGVIARSETEKMEES
                     VSEGISSPAVDTHTPDAPACKKKGKRRPFRPQPHWTHEPVEKVKSLPFHLRAEG
                     GGFLPLKIGTNDFAAFLTEHPETELTLDEWSCASCITTRQVLQRTAVAGIPRYGL
                     DGLPAGVQVSBCDALNARRWLAVRBQOKLKMKTMQEQSASTEKTER"
     CDS   complement(3130..4686)
                     /note="Orf5; similar to Salmonella typhimurium plasmid R64
                     IbfA, GenBank Accession Number BAB91598; growth inhibition
                     of bacteriophage BF23; contains one short conserved motive
                     of the family of ABC transporter proteins (pfam00005.5)"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42484.1"
                     /db_xref="GI:31075365"
                     /translation="MAEKNKKTITQVLNSIKINKLKCINGLNEIIFKPHALTAILGP
                     NGGGKSTILHAIASIYMPEEGFPGEHDHRLMHFFPRSPHAEWNGSDFIVNLTYRDGVM
                     IENELKNYGKADIRGSRWIQIYARPLREYVYLGIDKCVPIIESEKKNNIQYETESVS
                     NDLIINILHYASYILNKPYTSFNQHQQPNGKILIGVESGGLAYSSLSMSAGESQKIFLI
                     LETILKADKNALIIDEDLLIHDEALKKLIEVISSHAKDKNKQIIFTHRBMITTLS
                     DKINIRHVVNIIQGRSYSFEETKPDAINRLTGESTTPIEIVVEDDLAVAIINKICSSLK
                     ASKYVKIFKPGAASNAFTLLASTLIRGDNLSGKYILLDGDKYSTENEKKTALDKVFTG
                     TESRTYELKXAAASGKVKQFNLPNGVKPEQYIHYLITNVPLDGLGGGEYLEIIEARDIR
                     VELDAHNYISNILTKLGIDRPSGLITRVMDLASRPEWHQVSEVTDWLQPVVSDLMER
                     LPENDTVDIT"
     CDS   complement(4949..5539)
                     /note="Orf6; similar to Salmonella typhimurium plasmid
                     pR64 YdeA, GenBank Accession Number BAB91599"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42485.1"
                     /db_xref="GI:31075366"
                     /translation="MTQSRRPSPLQRRVLIVILAALDEKRPGPVLTRDLERVLERSGEA
                     PVYGPNLRASCRRLEDAGWLRTLRAPNLQLAVELTDAGRAVAQPLLLAEQDRLRAEQR
                     AABVVVLPLVPAAGLPADGTSATDLAVQLNGITYQACRGDFVVRLDGGSTCLQLMNKEG
                     RVRLEGDPLERVAQWLQACHDAGMEVRVQVNESVTP"
     CDS   complement(5539..5796)
                     /note="Orf7; similar to Salmonella typhimurium plasmid
                     pR64 YdfA, GenBank Accession Number BAB91600; contains one
                     short conserved motive of the family of ABC transporter
                     proteins (pfam0005.6, ABC_transporter)"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAP42486.1"
                     /db_xref="GI:31075367"
                     /translation="MDBRVFSLSYEQLTRFAEKRIRRCNLDSCQGAIYLCESAXAGAV
                     LIFWHELAINGYASMNAIRQELIDADFORLRNLIWPBDDR"
     CDS   6150..8288
                     /note="Orf8; similar to Salmonella typhimurium plasmid
                     pR64 YdfB GenBank Accession Number BAB91601; contains one

short conserved motive of the family of ABC transporter
proteins (pfam00005.6, ABC_transporter)"

```
CDS              /codon_start=1
                 /transl_table=11
                 /product="unknown"
                 /protein_id="AAP42487.1"
                 /db_xref="GI:31075368"
                 /translation="MDIBIRHCNNIVRAHITLTADKLNIKPAPNGTGKSTLSRAISCA
                 ARDDICQSLQALMPFRLRGENPDSTGPIVIGADGIGDVMCRNREYVSQFTGPDELISD
                 SFNILIRNQHAERERBIEEMTQKIRAVFPDNTELNSLIDHLQELSNAFRSTSGISR
                 SSTGMRGLSGGNKIHIPAGLENYQPYIRSRRVRRIDWQYRGLRFSPLSDQCCPPCI
                 GDITGKEAQIRQVRSEYDKSTIKNLTAIIRLVENLGNYLTESARERLLAITMLQNGPE
                 AEHIEYLVALLKRQTDDILTEKLTALRGINVRSLQEQONVRWIKTLTARLIDLQFPPDLQSE
                 LMQGITDRLNAALMDLINLAGPLQGKINRHRDSMIRLIAQHKTNINNPFLTVAGYKRV
                 DIAGEBQRKLRLRHIDFDGYVSGGSQHLSYGERNAPAIVLFMYBCLSKNPGLIILDD
                 PISSPDKNKKPAILEMLPRRASGECLKNRTVLMLTHIDVBPVIDTLKSVRLLFSNQVTA
                 SCLRLSAGVIEELPVNDGDIMTFMQICKSITASADCEEIIKLIYLRRYFEIVDERGDA
                 YQLLSNLFHRRVVPLDYREPAPAAGSGYPKMAPEKIQOALRDIREYVDSFDYPRLQALV
                 SSPDEIKNLYRRCRNGYBKLQVFRLLBLDQDHPVIRKFVNETYHIENEFICQLDPSRF
                 DLIPEYVIMBCDKLPLPPPANNQSSVARIA"
complement(8450..8851)
                 /note="Orf9; similar to Salmonella typhimurium plasmid
                 pR64 Mck, GenBank Accession Number BAB91602, and
                 Salmonella enterica serovar Dublin protein VagP, PIR
                 Accession Number S22686; involved in coordination of
                 plasmid replication with cell division"
                 /codon_start=1
                 /transl_table=11
                 /product="unknown"
                 /protein_id="AAP42488.1"
                 /db_xref="GI:31075369"
                 /translation="MLDTCICISPIMREQPEAVLKRLEQAVLRGHRIVVSAITYSEMRF
                 GATGPKASPRHVQLVDBFCARLDAILPWDRAAVDATTKIKVALRLAGTPIGPNDTAIA
                 GHAIAAGAILVTNNTKBFERVPDLVLEDWVK"

CDS    complement(8863..9093)
                 /note="Orf10; similar to Salmonella typhimurium plasmid
                 pR64 Kor, GenBank Accession Number BAB91603, and
                 Salmonella enterica serovar Dublin plasmid protein VagC,
                 Swiss-Prot Accession Number Q05459; involved in
                 coordination of plasmid replication with cell division"
                 /codon_start=1
                 /transl_table=11
                 /product="unknown"
                 /protein_id="AAP42489.1"
                 /db_xref="GI:31075370"
                 /translation="MRTVSIFKNGNNRAIRLPRDLPFEGVSELRIVREGDSIILRPVR
                 PTWGSPLEYEKADPDFMAEREDVVSDEGRFNL"

CDS    9389..9679
                 /note="Orf11; similar to Salmonella typhimurium plasmid
                 pR64 YdjA, GenBank Accession Number BAB91604, and
                 Escherichia coli plasmid F YebA, GenBank Accession Number
                 BAA97908"
                 /codon_start=1
                 /transl_table=11
                 /product="unknown"
                 /protein_id="AAP42496.1"
                 /db_xref="GI:31075377"
                 /translation="MSVKITGLDKMQQLKEVERATBALNGSYDVHFDANDPV9IENA
                 ICBAYSMVYERASGYAATNPMVSPLIEHMKRNLRQOILDRAEQORESGQDQN"
9669..10568
                 /note="Orf12; similar to Salmonella typhimurium plasmid
                 pR64 YdjA, GenBank Accession Number BAB91605, and
                 Escherichia coli plasmid F YebB, GenBank Accession Number
                 BAA97909"
```

```
Qy     94 CATTTCTTCGTTTCGGGGAATTGGACAAGAGAAAACTGTTGATGGCAGCCAAAATTGTGGC 153
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26215 CATTTCTTCGTTTCGGGGAATTGGACAAGAGAAAACTGTTGATGGCAGCCAAAATTGTGGC 26156

Qy    154 GGTGCAGAAAATGTGTTGTTAAAAACAGAAACTCAGCAGAAACATTCGTGTAAAATGGATTGCTCGGT 213
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26155 GGTGCAGAAAATGTGTTGTTAAAAACAGAAACTCAGCAGAAACATTCGTGTAAAATGGATTGCTCGGT 26096

Qy    214 TTTATCACTTTGGCATCTATACTCCCGCTCGGAAGCCCGGGTATATTGCTCACAATAG 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26095 TTTATCACTTTGGCATCTATACTCCCGCTCGGAAGCCCGGGTATATTGCTCACAATAG 26039
```

```
RESULT 9
AB016759/c
AB016759     Escherichia coli CFT073
LOCUS        AB016759    306358 bp    DNA    linear    BCT 24-DEC-2002
DEFINITION   Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION    AB016759 AB014075
VERSION      AB016759.1  GI:26107527
KEYWORDS     .
SOURCE       Escherichia coli CFT073
  ORGANISM   Escherichia coli CFT073
             Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
             Enterobacteriaceae; Escherichia.
REFERENCE    1  (bases 1 to 306358)
  AUTHORS    Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
             Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
             Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
             Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE      Extensive Mosiac Structure Revealed by the Complete Genome Sequence
             of Uropathogenic Escherichia coli
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED    12471157
REFERENCE    2  (bases 1 to 306358)
  AUTHORS    Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
             Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
             Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
             Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE      Direct Submission
  JOURNAL    Submitted (20-JUN-2002) Genetics Laboratory, University of
             Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1..306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83..3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83..3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /codon_start=1
                     /transl_table=11
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues 1 to 1209 of 1218 from GenPept.129 ;
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
                     /db_xref="GI:26107528"
                     /translation="MRDRSVALRLRNDSASGSYPTDGGGLMPANHTPTPAQSWIVRLA
                     RVCWERKKLSVIVVVASVBTILLAALTPLLTRQAVNDALAGNPARLPWLACGILLIAF
                     FDRIGNYVRRGYAGMLSLWVQHTLRGVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                     VHTLLQMCPVPLAVFTYVIAGIAVMLWMSPAMTLIIVVCLVVCLAITALRARRVPAQT
                     GMASDQLANLTEHHIREVLAQISVTKSCVAEMREFTHWLRQSRQIVRVIGAVISQAMP
                     GATMLALPVLGQIVLLCYGGWSVWHGRIDLGTFVPAPSFLAMLTGPTRVLASFIVIAQ
                     RTQASVERVPALIDTRSQMEDGTESINSQGVGLELENMSFDYHHGDRHILSDISFSLR
                     AGETVAVVGASGSGKSTLLMLARFYDPCSGKIWLNTSBGRQNLRLEALRRRVGI
                     VFBDAFLFAGTVARNIAYGHPQATADDIRRAAAAGASDFINALPKGFDSLLTERGTN
                     LSGGQRQRIALARALLITAPDVLILDDTISAVDAVTBAEINTALGRYADEGHMLLVIAR"
```

```
Qy     34 CAAAGGTTTACTGTTGTGGAAACAAACCGACAGCAGCGAGGTAACACCAAAGGAAACCATCACTCAT 93
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26275 CAAAGGTTTACTGTTGTGGAAACAAACCGACAGCAGCGAGGTAACACCAAAGGAAACCATCACTCAT 26216
```

Query Match   87.8%;   Score 237;   DB 1;   Length 31869;
Best Local Similarity 100.0%;   Pred. No. 2.3e-52;
Matches 237;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
                     /translation="RRSTLQLASRVVLDKGRMVDTGTGPAELEARCPAFRALMTGDSDFLATSHSNELWP
                     ABFATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVANMFVIAALLI
                     ALDSAAGVGVLILLQHGIDSGVAAGDMGTIGLCALLLACLIVIGWCSYSLQTVFAARA
                     ABSVQHSVRLRSPGHMLRLGLPWHBKHADSRLTRWTVDVDSLARFLQNGLAGAATSLV
                     TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAAQARLEIGKVNSTLQE
                     KVSGLRVQOSHGQOBLEGARLRALSERFRATRVRAQKYLAVPYPFLTCTEBASYAAVL
                     LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQOLSGIVDAWOQATASGKHIDELLATE
                     GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
                     TLIKLIAGLYFPTHGINIIRIGVQOMLDDASLTEYRROIGIVLDQDVALPSSDIAENIRYSR
                     P8ATNEDVBIASQRAGLYEMVCNLPQGRTPVVNGGADLSAGQRQLIALARAQLANAH
                     ILLLDBATSCLDRTSBERRLMSSLTDVVHAGKHSALIVAARLITTAQRCDLIAVIDKGLL
                     AEYGTHBQLLSAGGLYKTLWHDSVSSTALHRQHMKEFTPG"

gene            complement(3882..5045)
                     /gene="iroB"
                     /locus_tag="c1254"
CDS             complement(3882..5045)
                     /gene="iroB"
                     /locus_tag="c1254"
                     /codon_start=1
                     /transl_table=11
                     /note="Escherichia coli O157:H7 ortholog: z1190"
                     /product="Putative glucosyltransferase"
                     /protein_id="AAN79728.1"
                     /db_xref="GI:26107529"
                     /translation="MRLPDLDRQAERDPLMRILFVGHPPLYGLLYPVLSLAQAFRVNG
                     HRVLIASGQFAQKAARAGLVVFDAARGLDSRAGYRHHEAQRKKSNIGTQMGNFSFPS
                     EEMADHLVBPAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVVTR
                     RPDKRLLVSLGTYKPMVDGLDIIAMVMDSASEVDABIIIHISANARSDLRSLPSNVR
                     LVDWIPMGVFLNGADGFIHHGGAGNTLLTAIHAGIPQIVFGQGADRPVNARVVAERGCG
                     IIPGDVGLSSNMINAPLNNRSLRKASEVAAEMAAQPCPGEVAKSLITMVQKG"

gene            complement(5535..5834)
                     /locus_tag="c1255"
CDS             complement(5535..5834)
                     /locus_tag="c1255"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79729.1"
                     /db_xref="GI:26107530"
                     /translation="MKPTLLQTVENICLQLLFVLFLFTGTLIDVEALVSCSSSVAXNVQ
                     RTPASSVSSVSVYSASVSRTPAADIMLLSNNPLPAPGQVSFFGCSLAASGIHVTILB"

repeat_region   5867..6577
                     /insertion_seq="IS15A1A-like"
gene            5961..6470
                     /locus_tag="c1256"
CDS             5961..6470
                     /locus_tag="c1256"
                     /codon_start=1
                     /transl_table=11
                     /note="Residues 18 to 169 of 169 are 94.07 pct identical
                     to residues 1 to 152 of 152 from SwissProt:40 ;
                     >sp|Q57334|T200_SALTY Transposase for insertion sequence
                     element IS200"
                     /product="Hypothetical protein"
                     /protein_id="AAN79730.1"
                     /db_xref="GI:26107531"
                     /translation="MRSGNCKCQQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRO
                     VFPRKRRAIGCILRKLCEWKSVRIILEACCADHIHMLVBIPPRQWSVSGFMGYLYKGKS
                     SLMPYBQFGDLKPKYRNRBFWCRGYIVDTVGKYTAKIQDYIKHQLREDKMGRQLSIPY
                     PGSPPTGRK"

repeat_region   complement(6578..6940)
                     /note="incomplete"
                     /insertion_seq="IS4"
gene            complement(6558..6960)
                     /locus_tag="c1257"
CDS             complement(6558..6960)
                     /locus_tag="c1257"
                     /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                     residues 316 to 428 of 442 from MG1655 : b4278"
                     /function="IS, phage, Tn"
                     /codon_start=1
                     /transl_table=11
                     /product="Putative conserved protein"
                     /protein_id="AAN79731.1"
                     /translation="MPRQPDQGYREIKQTMQLSRLTLRSKPELVEQELWGVLLAYNL
                     VRYQMIKWABHLKGYWPNQLSFSBSCGMVWRMLMTLQGASPGRIPELMRDLASMGQLV
                     KLPTRREYAFPRVVKERP"

gene            7651..8013
                     /locus_tag="c1258"
CDS             7651..8013
                     /locus_tag="c1258"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79732.1"
                     /db_xref="GI:26107533"
                     /translation="MLGACFYSWGTSPYYGFPPWGAELVLMMWPEACFMLLQSSLYSL
                     LDGEQVLFSFWBLNKKQYEBGFBFYQIPSGNIVASCSSCGSGVSSNSSAHYARCTDFMRH
                     YSLNNYAACQLGROACLS"

gene            7835..8269
                     /locus_tag="c1259"
CDS             7835..8269
                     /locus_tag="c1259"
                     /codon_start=1
                     /transl_table=11
                     /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                     residues 104 to 242 of 258 from GenPept:129 ;
                     >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                     [Shigella flexneri]"
                     /product="Hypothetical protein"
                     /protein_id="AAN79733.1"
                     /db_xref="GI:26107534"
                     /translation="MKDLSFIRFFLAILLLLVPPALBFPVIHRHITPGALJTLCVITAL
                     IITLLASLGGRLVSVKCISBISFIRRHRIECMWAGFMIYFWTFSLIAGWYRKPQFKKEY
                     QTLHYBNARYYVLARYDGRLVLSQSYRSGSRKPVIINGHRL"

gene            8477..8553
                     /locus_tag="c5534"
tRNA            8477..8553
                     /locus_tag="c5534"
                     /product="tRNA-Arg"
                     /function="RNA; tRNA"

gene            complement(8832..9290)
                     /locus_tag="c1260"
CDS             complement(8832..9290)
                     /locus_tag="c1260"
                     /function="putative enzyme; Transposases Insertion
                     Sequence Associated"
                     /note="Escherichia coli O157:H7 ortholog: z4324"
                     /codon_start=1
                     /transl_table=11
                     /product="Putative Transposase"
                     /protein_id="AAN79734.1"
                     /db_xref="GI:26107535"
                     /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                     LERMVVIKAPIAVRMLGLRQEGISBBTQNDSCKKILTPTBWKLLWVKLEGKQLPSQTP
                     TLKWACLKLGRWHDSKRTCRPGWVVWMDGWPRLQDMVEGYPVMKSLDQHI"

gene            complement(9302..9667)
                     /locus_tag="c1261"
CDS             complement(9302..9667)
                     /locus_tag="c1261"
                     /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                     residues 59 to 93 of 316 from GenPept:129 ;
                     transposase [Shigella flexneri];
                     >gb|AAK18437.1|AF348706_126 (AF348706) putative
                     transposase [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79735.1"
                     /db_xref="GI:26107536"
                     /translation="MAERLGEIQKRVITVCDRBADIWHYLHYKVSHGQRFVVRTAQNS
```

Query Match   77.9%;   Score 210.4; DB 1;   Length 306358;
Best Local Similarity   93.2%;   Pred. No. 2.6e-45;
Matches 220;   Conservative   0;   Mismatches 16;   Indels   0;   Gaps   0;

RESULT 11

```
LOCUS       AR129411                 309 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 6 from patent US 6187321.
ACCESSION   AR129411
VERSION     AR129411.1  GI:14117308
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Nolan,L.K., Horne,S.M. and Robinson,M.
  TITLE     Avian E.coli 18S polypeptide
  JOURNAL   Patent: US 6187321-A 6 13-FEB-2001;
FEATURES             Location/Qualifiers
     source          1..309
                     /organism="unknown"
BASE COUNT       69 a      65 c      69 g      78 t
ORIGIN
```

```
Query Match              76.1%;  Score 205.6;  DB 6;  Length 309;
Best Local Similarity    91.9%;  Pred. No. 5.2e-44;
Matches   217;  Conservative   0;  Mismatches  19;  Indels   0;  Gaps   0;
```

```
Qy   34 CAAACGTTTACTGTTGGGAAACAAACCGACAGCAGTAACACCAAAGGAAACCATCACTCAT 93
        ||||||||||||||||||||||||||||  |||||||||||||  ||  ||||||||| |||
Db  220441 CAGACGTTTACTGTTGGGAAACAAACCTACAGCAGTAACCTACAGGAAAACCATCACCCAT 220382

Qy   94 CATTTCTTCGTTCGGGAATTGGACAAGAGAAACTCGTTGATGCAGCCAAAAATTTGTGGC 153
        ||||||||||||||||||||||| ||||| ||| ||||| |||||| ||||||||||||||
Db  220381 CATTTCTTCGTTCGGGAATTGGACAAGAGAAAACTCGTTGATGCAGCCAAAAATTTGTGGT 220322

Qy  154 GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAACAATTCGTTAAATGGATTGCTCGGT 213
        ||||||||||||||||||||||| ||||||||||| |||||||||||  ||||||||||||
Db  220321 GGTGCAGAAAATGTTGTTAAAACAGAAACCCAGCAAACAATTCGTTAAATGGATTGCTCGGT 220262

Qy  214 TTTATTCACTTTTTGGCATCTATACTCCGCTGGAAGCCCGGGTATATATTGCTCACAATA 269
        |||||||| ||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  220261 TTTATTCACTTTTAGGCATTTATACTCCGCTGGAAGCGCGGGTGTATTGCTCTATGCTCACAATA 220206
```

RESULT 12

```
LOCUS       BLAMLOM                  326 bp    DNA     linear   PHG 04-MAR-1994
DEFINITION  Bacteriophage lambda bor gene.
ACCESSION   X55792
VERSION     X55792.1  GI:288763
KEYWORDS    envelope protein.
SOURCE      Bacteriophage lambda
  ORGANISM  Bacteriophage lambda
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
            lambda-like viruses.
REFERENCE   1  (bases 1 to 326)
  AUTHORS   Barondess,J.J. and Beckwith,J.
  TITLE     A bacterial virulence determinant encoded by lysogenic coliphage
            lambda.
  JOURNAL   Nature 346 (6287), 871-874 (1990)
  MEDLINE   90363299
  PUBMED    2144037
FEATURES             Location/Qualifiers
     source          1..326
                     /organism="Bacteriophage lambda"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10710"
     gene            24..317
                     /gene="bor"
     CDS             24..317
                     /gene="bor"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="GI:288764"
                     /db_xref="SWISS-PROT:P26814"
                     /protein_id="CAA39317.1"
                     /translation="MKKMLLATALALLIINGCAQQTFVQNKPAAVAPKETITHHFFVS
                     GIGQKKTVDAAKICGGAENVVKTETQQTFVNGLLGPIILGIYTPLBARVYCSQ"
BASE COUNT      104 a      72 c      69 g      81 t
ORIGIN
```

```
Query Match              76.1%;  Score 205.6;  DB 7;  Length 326;
Best Local Similarity    91.9%;  Pred. No. 5.2e-44;
Matches   217;  Conservative   0;  Mismatches  19;  Indels   0;  Gaps   0;
```

```
Qy   34 CAAACGTTTACTGTTGGGAAACAAACCGACAGCAGTAACACCAAAGGAAACCATCACTCAT 93
        ||||||||||||||||||||||||||||  |||||||||||||  ||  ||||||||| |||
Db   73 CAGACGTTTACTGTTGGGAAACAAACCTACAGCAGTAACCTACAGGAAAACCATCACCCAT 132

Qy   94 CATTTCTTCGTTCGGGAATTGGACAAGAGAAACTCGTTGATGCAGCCAAAAATTTGTGGC 153
        ||||||||||||||||||||||||||||| ||| ||||| |||||| ||||||||||||||
Db  133 CATTTCTTCGTTCGGGAATTGGACAAGAGAAAACTCGTTGATGCAGCCAAAAATTTGTGGT 192

Qy  154 GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAACAATTCGTTAAATGGATTGCTCGGT 213
        ||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db  193 GGCGCAGAAAATGTTGTTAAAACACTGTGATGCAAGCCAAAAACCATCACTCAT 252

Qy  214 TTTATTCACTTTTTGGCATCTATACTCCGCTGGAAGCCCGGGTATATATTGCTCACAATA 269
        |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  253 TTTATTCACTTTTAGGCATTTATACTCCGCTGGAAGCGCGGGTGTATTGCTCTATGCTCACAATA 308
```

```
RESULT 13

LOCUS       U02453                 9170 bp    DNA     linear   SYN 29-MAR-1996
DEFINITION  Cloning vector lambda EMBL3, right arm.
ACCESSION   U02453
VERSION     U02453.1  GI:413819
KEYWORDS    .
SOURCE      Cloning vector lambda EMBL3
  ORGANISM  Cloning vector lambda EMBL3
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 9170)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9170)
  AUTHORS   Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N.
  TITLE     Lambda replacement vectors carrying polylinker sequences
  JOURNAL   J. Mol. Biol. 170 (4), 827-842 (1983)
   PUBMED   6315951
REFERENCE   3  (bases 1 to 9170)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     This sequence has been compiled from information in the sequence
            databases, published literature and other sources. If you suspect
            there is an error in this sequence, please contact CLONTECH'S
            Technical Service Department at (415) 424-8222 or (800) 662-2566,
            extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES             Location/Qualifiers
     source          1..9170
                     /organism="Cloning vector lambda EMBL3"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31783"
BASE COUNT     2534 a   1996 c   2251 g   2389 t
ORIGIN
```

```
Query Match     76.1%;  Score 205.6;  DB 12;  Length 9170;
Best Local Similarity   91.9%;  Pred. No. 5e-44;
Matches  217;  Conservative   0;  Mismatches  19;  Indels   0;  Gaps   0;

Qy         141 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAGAAGAAAACTGTCGATGCAGCCAAAAATTTGTGGC 200
               ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         154 GGTGCAGAAAATGTTGTTAAAACAGAAAACTCAGCAGAAGAAAACTGTCGATGCAGCCAAAAATTTGTGGC 213

Qy         201 GGCGCAGAAAAATGTTGTTAAAACAGAAACCCAGCAAACATTCGTATAATTGCTCGGGT 260
               ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         214 GGCGCAGAAAATGTTGTTAAAACAGAAACCCAGCAAACATTCGTATAATTGCTCGGGT 269

Qy         261 TTTATCACTTTTGGCATCTATACTCCGCTGGAAGCGCGTGTGTAATTGCTCACAATA 316
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 TTTATCACTTTTAGGCATTTATACTCCGCTGGAAGCGCGTGTGTAATTGCTCACAATA 316
```

```
RESULT 14
XXU02427/c
```

```
LOCUS       XXU02427               9205 bp    DNA     linear   SYN 29-JAN-1997
DEFINITION  Cloning vector lambda EMBL3 SP6/T7, right arm.
ACCESSION   U02427
VERSION     U02427.1  GI:413793
KEYWORDS    .
SOURCE      Cloning vector lambda EMBL3 SP6/T7
  ORGANISM  Cloning vector lambda EMBL3 SP6/T7
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 9205)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9205)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     This sequence has been compiled from information in the sequence
            databases, published literature and other sources. If you suspect
            there is an error in this sequence, please contact CLONTECH'S
            Technical Service Department at (415) 424-8222 or (800) 662-2566,
            extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES             Location/Qualifiers
     source          1..9205
                     /organism="Cloning vector lambda EMBL3 SP6/T7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31784"
BASE COUNT     2546 a   2002 c   2253 g   2404 t
ORIGIN
```

```
Query Match     76.1%;  Score 205.6;  DB 12;  Length 9205;
Best Local Similarity   91.9%;  Pred. No. 5e-44;
Matches  217;  Conservative   0;  Mismatches  19;  Indels   0;  Gaps   0;

Qy          34 CAAAGCGTTTACTGTTGGAAACAAAGCAGCGAGTAACACACACACCAAAGGAAAACCATCACTCAT 93
               ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        7398 CAGACGTTTACTGTTCAAAACAGAAAACCGGCACACAGTAGCACCAAAGGAAAACCATCACTCAT 7339

Qy          94 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAAAACTGTTGATGCAGCCAAAAATTTGTGGC 153
               ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        7338 CATTTCTTCGTTTCGGGAATTGGACAAGAGAAAAAACTGTTGATGCAGCCAAAAATTTGTGGC 7279

Qy         154 GGTGCAGAAAATGTTGTTAAAACAGAAACTCAGCAGAAGAAAACTGTCGATGCAGCCAAAAATTTGTGGC 213
               ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        7278 GGCGCAGAAAAATGTTGTTAAAACAGAAAACCCAGCAAACATTCGTAAATGGATTGCTCGGGT 7219

Qy         214 TTTATCACTTTTGGCATCTATACTCCGCTGGAAGCGCGTGTGTAATTGCTCACAATA 269
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7218 TTTATCACTTTTAGGCATTTATACTCCGCTGGAAGCGCGTGTGTAATTGCTCACAATA 7163
```

```
RESULT 15
AE005297/c

LOCUS       AE005297              11313 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 121
            of 155.
ACCESSION   AE005297
VERSION     AE005297.1  GI:12514322
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 11313)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
   MEDLINE  21074935
```

```
PUBMED    11206551
REFERENCE   2  (bases 1 to 11313)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1..11313
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1..>11313
                     /note="O-island #45; Region of the EDL933 chromosome not
                     homologous to B. coli K-12 MG1655"
     misc_feature    <1..>11313
                     /note="Prophage BP-933W; includes one copy of the 7 bp
                     direct repeat that flanks the prophage"
     gene            69..284
                     /gene="Z1468"
     CDS             69..284
                     /gene="Z1468"
                     /function="putative membrane; Lysis (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 71 of 71 are 100.00 pct identical to
                     residues 1 to 71 of 71 from GenPept 118 :
                     gi|4585421|gb|AAD25449.1|AF125520_44 (AF125520) protein S
                     [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lysis protein S of bacteriophage
                     BP-933W"
                     /protein_id="AAG55591.1"
                     /db_xref="GI:12514323"
                     /translation="MYQMEKITTGVSYTTSAVGTGYWFPLCLLDRVSPSQWAAIGVLGS
                     LLFGLLTYLTNLYFKIKEDRKAARGE"
     gene            289..822
                     /gene="Z1469"
     CDS             289..822
                     /gene="Z1469"
                     /function="putative enzyme; Lysis (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from GenPept 118 :
                     gi|4585422|gb|AAD25450.1|AF125520_45 (AF125520) endolysin
                     [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lysozyme protein R of bacteriophage
                     BP-933W"
                     /protein_id="AAG55592.1"
                     /db_xref="GI:12514324"
                     /translation="MSRKLRYGLSAAVLALIAAGASAPEILDQFLDEKEGNHTTAYRD
                     GAGIWTICRGATRVDGKPVIPGMKLSKEKCDRVNAIERDKALAWEKNIKVPLTEPQK
                     AGIASPCPYNIGPGKCPFPSTPYRRINAGDRKGACEAIRWIKDGGRDCRIRSNNCYGO
                     VSRRDESALACWGIDR"
     gene            1093..1662
                     /gene="Z1471"
     CDS             1093..1662
                     /gene="Z1471"
                     /function="putative regulator; Lysis (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="putative antirepressor protein Ant of
                     bacteriophage BP-933W"
                     /protein_id="AAG55593.1"
                     /db_xref="GI:12514325"
                     /translation="MNMMAVPFHGNSLIYVNHNGEPYVPMKPVVAGMGLAWGSQLAKL
                     RQRPASTIREIVMVABDGRQRNMVSMPLRKCLAGHHLQTINPNKVKPEIRDKVIRYQBEC
                     DDVLKBYWTKGFVVNPRKGMSVMEELNQACADMKRDXNIASVFATGLNEWKQVKAAHVS
                     KIRTLVNRANMLIDPVLADTGKGKITKAD"
     gene            1816..2280
                     /gene="Z1473"
     CDS             1816..2280
                     /gene="Z1473"
                     /function="putative enzyme; Lysis (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 154 of 154 are 100.00 pct identical
                     to residues 1 to 154 of 154 from GenPept 118 :
                     gi|4585424|gb|AAD25452.1|AF125520_47 (AF125520)
                     endopeptidase Rz [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative endopeptidase Rz of bacteriophage
                     BP-933W"
                     /protein_id="AAG55594.1"
                     /db_xref="GI:12514326"
                     /translation="MNRVLCVIIVLAVGVGALWLATNHYRDNALTYKAQRDKKAREL
                     EQANATITDMQVRQRDVAALDAKYSRELADARAENETLRADVAAGRKRLRINATCSGT
                     VREATCTSGVDNATGPRLAPTAERDYFILRERLITMQKOLEGTQKYINBQCR"
     gene            complement(2312..2605)
                     /gene="borW"
     CDS             complement(2312..2605)
                     /gene="borW"
                     /function="putative Bor protein precursor of bacteriophage
                     BP-933W"
                     /note="synonym: Z1474"
                     /codon_start=1
                     /transl_table=11
                     /product="putative Bor protein precursor of bacteriophage
                     BP-933W"
                     /protein_id="AAG55595.1"
                     /db_xref="GI:12514327"
                     /translation="MKKMLLATPALALLIITGCAQQTFTVQNKQTAVAPKETITHHFPVS
                     GIGQKKTVDAAXKICGGTENVXNTETQQTFVNGLLGFITLGIYTPLERARVYCSQ"
     gene            3014..3820
                     /gene="Z1475"
     CDS             3014..3820
                     /gene="Z1475"
                     /function="putative enzyme; DNA packaging, phage assembly
                     (Phage or Prophage Related)"
                     /note="Residues 1 to 268 of 268 are 100.00 pct identical
                     to residues 1 to 268 of 268 from GenPept 118 :
                     gi|4585427|gb|AAD25455.1|AF125520_50 (AF125520) putative
                     small subunit terminase [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative terminase small subunit of
                     bacteriophage BP-933W"
                     /protein_id="AAG55596.1"
                     /db_xref="GI:12514328"
                     /translation="MAKGLDWKLEQAPRRHAHTGITLLDWCRKKKINYNTARTRIKM
                     GKIDHBIDHKTDHEIDHDISDEPFCNDAGSGDBKCAKNSEKQCANSABTKRIRGSRLL
                     PPSNAPSQRNTHAVRHRGYAXYLBADNIMDDASDMVLFDELVFTRARALSVTKALKGM
                     FADLRBRATDVSTRVALYDKILKAEAQALDRNIARIESTERSLLTLDVLABTAPKGLRADR
                     ERINAARDKLRAFTDILTNQRRGVVTPVSDIVSSLHEMSNSGRLDDIPEE"
     gene            3801..5507
                     /gene="Z1476"
     CDS             3801..5507
                     /gene="Z1476"
                     /function="putative enzyme; DNA packaging, phage assembly
                     (Phage or Prophage Related)"
                     /note="Residues 1 to 568 of 568 are 100.00 pct identical
```

```
            to residues 1 to 568 of 568 from GenPept 118 :
            gi|4585428|gb|AAD25456.1|AF125520_51 (AF125520) putative
            large subunit terminase [Bacteriophage 933W]"
            /codon_start=1
            /transl_table=11
            /product="partial putative terminase large subunit of
            bacteriophage BP-933W"
            /protein_id="AAG55597.1"
            /db_xref="GI:12514329"
            /translation="MTFRKNEPRCDEPSEMTEAEQRLPIMTKLSNPWWRLNHLYKIQN
            EKGELVTFRNRPAQKQLFRSMEHRKNIILKARQLGFSTAIDIYLLDQALFIPHLKCGIV
            AQDKQAASIFRTKIAVPFDHLIPDWLRAGFTIVSRRSGRASGGYILFGRGSSIQVATSF
            RSGTVQRLHISEHGKICAKYPAKAKELRTGTLNAVSDECIIPDESTAEGVGGDYEMS
            NRAQEITAGGLLLTAQDYKFHFYAWWQDPKYSARVPESGLKLSRSKMTYFSAVEKAMN
            ITLTDBQKQWYINKETBQRREMKQFPSTPQEAPLITSGRRVFSAESTLQARSPCSPPM
            IVVDIEBPVTGAKTKAQSLREGNKNVELQRTLMNYILLVWELPDPDEBYVCGADTAEGLEH
            GDRSBLDVVRSNGEQVAHWFGHLDARLFAHLISQVCRMTNNAFVGPERNNHGHAVIL
            KLRRLYFTRYIYNEQHLDQAYDDDTPRLGWLTTRQSKPVLTBGMKTLLNNGISGIRWS
            GTLSBMNTVYDAKGSMNAQBGCPDDQLMSYMIAQBMRARMPVRVKQKTDKRRTHWM
            AH"
gene        5507..7651
            /gene="Z1477"
CDS         5507..7651
            /gene="Z1477"
            /function="putative structure; Structural component (phage
            or Prophage Related)"
            /note="Residues 1 to 714 of 714 are 100.00 pct identical
            to residues 1 to 714 of 714 from GenPept 118 :
            gi|4585429|gb|AAD25457.1|AF125520_52 (AF125520) putative
            portal protein [Bacteriophage 933W]"
            /codon_start=1
            /transl_table=11
            /product="putative portal protein of bacteriophage
            BP-933W"
```

```
Query Match     76.1%;  Score 205.6;  DB 1;  Length 11313;
Best Local Similarity   91.9%;  Pred. No. 5e-44;
Matches 217;  Conservative   0;  Mismatches 19;  Indels   0;  Gaps   0;

Qy     34 CAAACGTTTACTGTTGGGAAACAAACCGACACGCAGTAACACCAAAGGAAACCATCACTCAT   93
          |||||||||||||||||||| |||||||||| |||||||||||||||||||||||||||||
Db   2548 CAGACGTTTACTGTTCAAAACAAACACGCAGTAACACCAAAGGAAACCATCACCCAT   2489

Qy     94 CATTTCTTCGTTTCGGGAATTGGCACAGAGAAAACTGTTGATCGAGCCAAAATTGTGGC   153
          |||||| |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   2488 CATTTCTTCGTTTCTGGAATTGGGCAGAGAAAACTGTCGATGCAGCCAAAATTTGTGGC   2429

Qy    154 GGTGCAGAAAAATGTTGTTAAAACAGAAACTCAGCAAACATTGGTAAANGGATTGCTCGGT   213
          |||||||||| ||||||||||| | ||||||||||||||| ||||| ||||||||||||
Db   2428 GGCACAGAAAATGTTGTTAAAACAGAAACCCAACATTCGTAAAATGGATTGCTCGGT   2369

Qy    214 TTTATCACTTTTGGCATCTAACTCCGCTGGAAGCCCGGGTATATTGCTCACAATA   269
          ||||| ||||||||| |||| |||||||||||||||||||||||||||||||||
Db   2368 TTTATTACTTTTAGGCATTAACTCCCTGGAAGCGCGGGGTGTATTCGCTCACAATA   2313
```

Search completed: January 21, 2004, 02:44:39
Job time : 1539 secs